Physiological Sensing based Stress Analysis during Assessment

Aniruddha Sinha, Pratyusha Das, Rahul Gavas, Debatri Chatterjee
TCS Research, Innovation Lab,
Tata Consultancy Services Ltd.
Kolkata, India
Email:{aniruddha.s, das.pratyusha, rahul.gavas, debatri.chatterjee}@tcs.com

Sanjoy Kumar Saha
Computer Science & Engineering
Jadavpur University
Kolkata, India
Email: sks_ju@yahoo.co.in

***Abstract*—During an examination, the performance of a student not only depends on the preparation but also depends on cognitive and psychological factors. The present study aims at analysing the dynamics of the mental workload or stress and its impact on students' performance during an examination process. The experiment is designed using a set of multiple choice questions (MCQs) in one of the subject area namely data-structure. The MCQs are selected in three levels of complexity, which are rated by subject matter experts. A total of 13 right-handed graduate students in the specialization of Masters in Computer Application participated in the experiment. In order to get rid of any bias related to the intelligent quotient (IQ), students are taken such that the IQ varies from low to high. Experimental tasks are designed as interleaved sequences of the MCQs of varying complexity levels. During the tasks, Galvanic skin response (GSR) and pulse-oximeter (SPO2) signals are captured to estimate the stress of an individual from electro-dermal changes and heart rate variability (HRV). Apart from the tonic and phasic power of the GSR, fluctuation analysis is performed on the raw GSR signals. The standard deviation of normal-to-normal interval (SDNN) is computed to measure the HRV using the Photoplethysmogram (PPG) signal obtained from the SPO2 device. An estimation of the stress level is done using a score computed by the fluctuation analysis of the GSR signal. Out of the 13 students, the GSR sensor data for the 6 students are found to erroneous and hence only 7 data are analyzed. The methodology used to reject the erroneous data is also presented. Results indicate that there is a strong statistical correlation between the complexity of the MCQs and the GSR signals. Results of HRV analysis provide new insights to guess work during an assessment process.**

Keywords— Multiple choice; Undergraduate; Behavior theory; GSR; SPO2; PPG; Stress

I. INTRODUCTION

The academic performance of an individual not only depends on the amount of study and preparation an individual has done but also on the complex outcome of cognitive and psychological factors. Hence, during the evaluation, it is very important to understand how the state of an individual affects the performance. This would enable the evaluation system to get new insights and provide constructive feedback to an individual for future improvements.

In conventional education scenarios, the learning content provider basically assesses the learner, based on the performance in the assessment. This is a widely practiced and well established scheme. The tutor tries to get an insight of the learner's capabilities based on the score obtained at the end of teaching a particular module. However, this technique does not give a complete insight of the knowledge gained or the efforts put by the student. Hence, there is a need for a new technique to monitor these effects.

Approaches to monitor stress or the cognitive load via questionnaire based approaches are widely practiced. Popular questionnaires like the NASA TLX [1], PAAS [2], the Bratfisch-Borg-Dornic's scale [3], have been widely used to analyze stress in the form of cognitive load imposed while performing in various learning based tasks. However, monitoring of the learner throughout the process is not possible or convenient with the questionnaire based approaches. This creates an urge to adopt alternative solutions like sensor technologies to serve the purpose.

The cognitive and emotional states are governed by the challenge of a given task and the skill level of an individual. A mismatch between the skill and the challenge leads to anxiety or boredom as shown in Fig. 1. People become bored [4] whenever the skill is much higher than the task complexity whereas, in case of lower skill an individual becomes anxious leading to stress [4]. When the skill and challenge are roughly proportional, people enter *Flow state* i.e. a state of focused concentration and enjoyment.

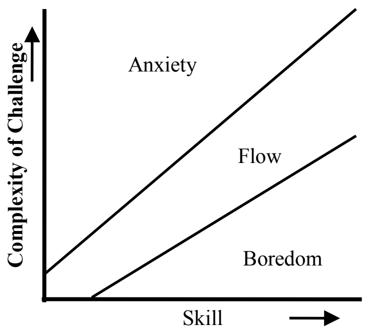

Fig. 1. Boredom, Flow and Anxiety state for skill and challenge

This study mainly aims at extracting learner's feedback in the form of stress when subjected to learning assessments using low cost, user friendly, low resolution sensor devices. The targeted sensors in this study are Galvanic Skin Response

(GSR) and pulse–oximeter (SPO2). The aim is to use useful features to derive stress based scores which are expected to help the tutor to understand the stress levels that the learner undergoes during assessment based scenarios.

So far very few attempts for direct measurement of cognitive load or stress have been seen. Most of current approaches use indirect methods for measuring the anxiety and flow [5], [6], [4]. In this paper we have used commercially available GSR device from Shimmer[1] and SPO2 from Contec[2]. The sampling frequencies used for the GSR and SPO2 devices are 10.2 Hz is 60 Hz respectively. There are several approaches to study cognitive load using brain activations like functional Magnetic Resonance Imaging (fMRI) [6], functional Near Infra-Red (fNIR) [7] etc. However, these are costly and require specialized setup. Hence we have used two physiological sensing namely, heart rate variability (HRV) and Galvanic Skin Response (GSR) for measuring the level of stress in mental workload during an assessment process.

In this paper, the approach involves designing a set of MCQs in the area of data-structure which are categorized into three different levels of complexity namely, easy, moderate and difficult. The levels of the MCQs are decided based on the rating from the experts of the subject area. A set of MCQs are then presented to the students with the GSR sensor worn on the medial phalanges of the index and middle finger and SPO2 device worn on the ring fingers of the left hand. The sensor signals are captured in synchronous with the response to the MCQs. The GSR and SPO2 signals are then analyzed to correlate with the level of the complexity of the MCQs. Additionally, feedback is taken from the students against each MCQ, on whether the answer provided is by guess work or by the required mental effort. This is used to analyze the change in HRV due to the change in the complexity levels of the MCQs for a given amount of guess work.

The paper is organized as follows. Section II reviews some of the closely related work in the domain of analyzing stress levels using sensors. Section III briefs the methodology employed in the experimentation. The results are discussed in Section IV. The paper concludes in Section V along with the future roadmap of the proposed approach.

II. RELATED WORK

The focus towards sensor based assessment of cognitive states like stress is currently getting research importance. Usage of Electroencephalogram (EEG), Heart rate, pupil diameter, GSR, heat flux to monitor physiological changes during involvement in stress inducing psychological tasks have been done as in [8]. Most of the existing works monitors the stress levels mainly for standard psychometric tests. The sensor based technology seems to be very conducive for such cases. However, the challenge lies in using the same in case of real life scenarios. This mainly involves education oriented cases like learning, assessment, etc. Recently, training programs to evaluate attention [9], flow state (or the state of total involvement) [10] are being designed using sensors. Stress based mental states get more evident in the GSR sensor, as an increase in stress is positively correlated with increase in sweating. GSR is a sensitive and convenient measure of changes in sympathetic arousal associated with emotion, cognition, and attention [11]. This effect is very well seen in the heart rate (HR) data as studied in [12]. GSR has been used in applications like driver stress monitoring [13], affective computing based projects [14] and providing auditory stimuli [15]. Other real time applications of stress detection using GSR and HR is driver behavior modeling [16], human-computer interactions (HCI) [17], security applications [18], clinical diagnosis [19], etc.

Analysis of GSR signals involve a mixture of various time and frequency domain features. The raw GSR signal is shown to be an indicator of stress by Lin et al. [20]. For arithmetic and reading tasks, normalized GSR signal is used by Nourbakhsh et al. [21]. Whereas, time domain features like peak magnitude, peak duration and peak area are used by Zhou et al. in [22]. They have derived the frequency domain features by Z-score normalization and used a time consuming supervised approach for classification. Tonic power of the GSR signals is shown to be a good indicator of stress [10]. Mancilla et al. [23] have used GSR, skin temperature and heart rate as physiological parameters to quantify the stress in the range of 0 to 100. However, all these experiments are done using costly GSR setups with high sampling rate which is a hindrance for mass deployment.

Given the above landscape, it can be seen that very less work is done on understanding the effect of the complexity of the MCQ on the cognitive load and stress level. Moreover, there is no work on quantifying the stress using physiological sensing for a given set of MCQs.

III. METHODOLOGY

The present study mainly focuses on modeling the physiological changes during an assessment. The stimulus used for the experiment is designed to replicate an online examination process based on data structures. Students give an MCQ based examination during which the GSR and SPO2 signals are captured to analyze the cognitive stress as depicted in Fig. 2.

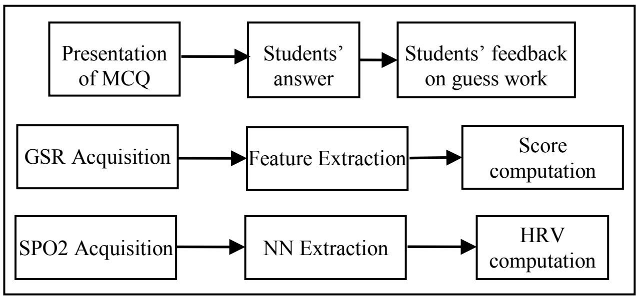

Fig. 2. Overall methodology – presentation of stimulus (MCQ), acquisition and processing of GSR and SPO2 signals

A. Design of MCQs

Initially 40 MCQs are taken from a question bank of data structure. These MCQs are rated by three experts in the subject matter who are the faculty of Computer Science in an Engineering Institute. The rating is a number between 1, 2 and

[1] http://www.shimmersensing.com/shop/wireless-gsr-sensor

[2] http://www.contecmed.com/index.php?option=com_virtuemart&Itemid=592

3 indicating *easy*, *moderate* and *difficult* respectively. Majority voting is used to select the final complexity of the MCQ. In order to limit the duration of the experiment, six *difficult*, eight *moderate* and twelve *easy* MCQs are selected. In order to maintain the duration of the stimulus for each complexity level to a similar value, the number of MCQs are increased with decrease in the complexity level. This is due to the fact that more duration is expected to be taken by the students for MCQs with higher complexity level. In order to analyze the significance of the difference of complexity, in terms of the ratings, one way ANOVA analysis is performed on the ratings given to the selected MCQs. The F-value is found to be 49.55 with p-value of 2.93e-14, thus justifying the statistical significance in the separation of the complexity levels of three types of MCQs.

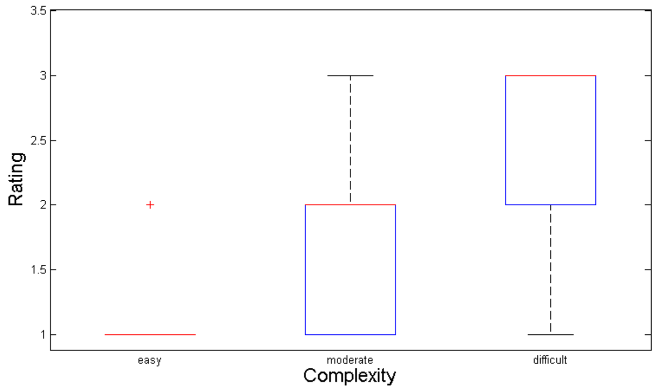

Fig. 3. Output of oneway ANOVA analysis of the ratings provided to three complexity levels (*easy*, *moderate* and *difficult*) of MCQs by the experts.

B. *Presentation of the Stimulus (MCQ)*

The stimulus for the experimentation is a set of MCQs. The stimulus is implemented using Python and is displayed on a standard computer monitor with a resolution of 1366x768 pixels. The stimulus is divided into 3 categories - easy (12 questions), moderate (8 questions) and difficult (6 questions) to induce 3 levels of cognitive loads namely, low, medium and high loads, respectively. The students had to click on the option containing the right answer and proceed to the next one. The options are aligned to the bottom left portion of the screen as shown in Fig. 4. Once the student selects an option then the "Next Slide" button is clicked to move to the next MCQ. Between the two MCQs, a feedback slide is presented to the student where a feedback is taken whether the answer is provided by guess or not. This is to get inputs on the actual mental workload that a student experienced during answering the MCQ.

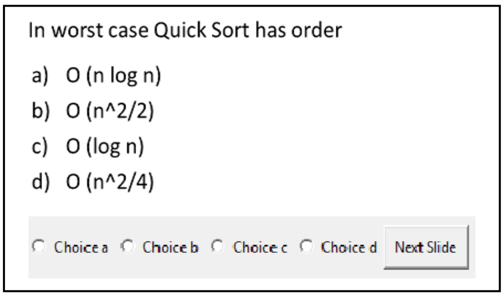

Fig. 4. Screen shot of the presentation of the MCQ

The sequence of tasks that the participants are subjected to is shown in Fig. 5. Before the onset of every stimulus, the student is asked to relax with eyes closed condition for a period of 2 minutes. This is followed with a preparation time of 5 seconds and then a baseline of 5 seconds is provided. After that, the presentation of MCQ stimulus is started.

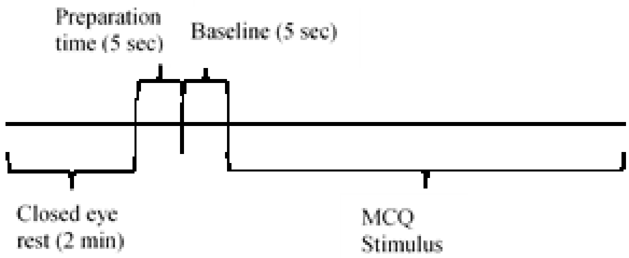

Fig. 5. Sequence of tasks and presentation of stimulus

C. *GSR Data Analysis*

The GSR device applies a constant voltage to the skin through two electrodes. The voltage is so small that it cannot be felt or perceived by the individual. However, the current that flows through the skin, as the voltage is applied, can be detected by the device. The change in current, provides the instantaneous change in skin conductance and reflects a physiological response. Though the pattern of various mental states especially arousal is evident in the GSR data however, the process of data collection is not easy as the GSR data consists of noise due to loose contacts with the skin, environmental temperature (cold/hot), activity (exercising), and personal differences. These factors are a major source of artifacts in the data. Along with this, there are other issues like lability and stability of the participants with respect to the GSR data [24]. Hence, the interpretation of stress levels from the raw GSR series is not straight forward [25]. To overcome these factors, we first did an initial screening to detect and avoid labiles and stabiles in the study. This was done while ensuring a conducive normal room temperature in the data acquisition room.

1) *Feature Extraction*

In order to understand the effect of mental stress have we calculated following features of the GSR signal as explained below.

a) *Tonic and Phasic Power*

The GSR signal is characterized by two components: a fast component called 'phasic' and a slow component called 'tonic'. Both tonic and phasic components contain information associated with specific physiological aspects of brain states. Here the tonic component is calculated only taking the inverse transform of first few Fourier coefficients as in (1), whereas the phasic component is calculated by inversing the higher coefficient of Fourier coefficients as given in (2).

$$tonic\ component = IFFT(\sum_{n=0}^{N-1} x(n).e^{-j(\frac{2\pi}{N})nk})\ ,\ k=0,1,2,3 \qquad (1)$$

$$phasic\ component = IFFT(\sum_{n=0}^{N-1} x(n).e^{-j(\frac{2\pi}{N})nk})\ , k=4,5,...N-1 \qquad (2)$$

b) Peak detection algorithm

A significant peak [26] in GSR signal is considered as a peak height greater than $0.05\mu S$ from its previous trough. In order to remove high frequency artifacts, the raw signal $x(t)$, sampled at a frequency of 10.2 Hz, is passed through a 2 Hz low-pass filter. The difference of successive input samples $y(t)$ is obtained using (3).

$$y(t) = x(t+1) - x(t) \tag{3}$$

A peak at $t = t_p$ is considered, if $y(t_p) < 0$ and $y(t_p - 1) < 0$. Similarly, a valley is considered at $t = t_v$, if $y(t_v - 1) < 0$ and $y(t) > 0 \; \forall \, t = t_v$ to $(t_p - 1)$. The difference between the peak and the valley is the height *h* (4). It is considered as a significant peak for $h > 0.05\mu S$.

$$h = x(t_p) - x(t_v) \tag{4}$$

c) Fluctuation analysis of GSR data

A fluctuation index is derived of a signal $x(t)$ of length *L*. Initially the zero mean signal $x_{zm}(t)$ is computed by subtracting the mean μ_x of the signal, as given by (5).

$$x_{zm}(t) = x(t) - \mu_x, \forall t \tag{5}$$

Then the signal is divided into small windows of length say *l*. A local trend of the signal is computed by fitting a line in each window while minimizing the least square error. The local trend is a time-series $x_l(t)$, computed for all the windows, over the entire length *L*. Finally, the Fluctuation index *(FI)* is calculated as given in (6).

$$FI = \sqrt{\frac{1}{L}\sum_{t=1}^{L}[x_{zm}(t) - x_l(t)]^2} \tag{6}$$

d) Computation of Score

A normalized score for the mental effort is derived from the features of GSR. During performance a task, initially the effort level increases substantially due to the anticipation/excitement of the task itself. Individuals tend to put more effort with increase in the difficulty level of the task. Finally, the effort reaches a saturation point as the difficulty level is further increased.

The window-wise GSR features are computed for both eye closed rest condition and while performing the task. The minimum value of the feature in eye closed condition is considered as $f(R)$. During the task condition, the feature value at a window (w) is $f\ (w)$. A score for the Effort Index (*EI*) as shown in (7) is derived from $f(R)$ and $f\ (w)$.

$$EI(w) = 1 - \frac{f(R)}{f\ (w)} \tag{7}$$

where $f\ (w) > 0$, $f(w) \geq f(R)$ and $0 \leq EI(w) \leq 1$.

A non-linear relationship between $EI(w)$ and $f\ (w)$ can be seen in Fig. 6. A sharp rise of $EI(w)$ can be observed at the low level task, where the $f\ (w)$ is low. However, as the difficulty level of the task increases, the score reaches a saturation with $EI \approx 1$.

This score in addition to other existing measures like performance score, completion time etc., gives an additional information and teachers can have more insight into the actual mental state of the student and thus helps in better evaluation of student.

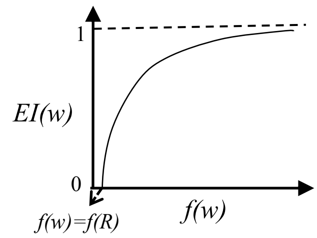

Fig. 6. Relationship of $EI(w)$ and $f(w)$

D. SPO2 Data Analysis

Pulse oximeter gives a measure of oxygen saturation in the blood. We have used CMS-50D plus oximeter from Contec. It comes with a software for measuring and recording the oxygen saturation and pulse rate data accurately. It also provides the Photoplethysmogram (PPG) data, from the finger tip, as a time series information at 60 Hz sampling rate. The PPG is a pulsating signal which contains the blood flow information due to systolic and diastolic cycles of heart. The device is very light weight and the performance does not depend upon the ambient light.

We have used the SPO2 device wearable on the ring finder for sensing the PPG to measure the heart rate (HR) and the heart rate variability (HRV). The HRV is a measure of the variations in instantaneous HR which is derived using the peak to peak (R-R interval) time difference in the PPG signal. The most recommended measure of HRV is SDNN (standard deviation of the normal-to-normal (n-n) heart beat) [28]. The level of stress is reflected in the HRV [29]. The nervous system responds to stress and controls the HRV. When a subject is in relaxed state, the HRV is higher compared to that under stress. The reason is, under stress condition, the parasympathetic nervous system responds and lowers the HRV. In the present experiment, the HRV is analyzed to understand the level of stress in the participants while they undergo the MCQ based assessment. We have calculated SDNN (Successive difference between NN Intervals) which is the most important time domain HRV parameters [28].

IV. RESULTS AND DISCUSSIONS

This section provides the results for the MCQ based experiment. Analysis is done for both the GSR and the SPO2 data for three levels of complexities in MCQs. Initially the description of the participants are given and the methodology used to screen them is provided. Once the sensor data are captured, they are manually observed to verify the correctness of the data. This is important to reject the noisy data due to loose contact and wireless communication error. The clean data is used for processing and the results are presented for the same.

A. Experiment Protocol

At the beginning of the experiment, the IQ levels of the participants are measured by a standardized IQ[3] test prior to presentation of the stimulus. The test consists of 20 mathematical series questions. The IQ score that is generated based on the test is not shared with the participant and is only used to make sure that the group of participants that we have selected has a good amount of spread in the IQ level. The scores of IQ less than 89 are treated as low IQ, scores between 90 and 109 are treated as Medium IQ and scores with above 110 is treated as High IQ.

Once the IQ tests are over participants are randomly selected to form two groups with the constraint of having balanced male and female ratio and IQ levels.

One of the group is presented with the stimulus (MCQs) sequence of *easy*, *moderate* and *difficult* and the other group is presented with the reverse sequence. All the MCQs for each of the complexity level are presented together. A time separation of 5 minutes is given between two levels of MCQs. In each level of MCQs, after the participants answer an MCQ, an immediate feedback is taken on whether the answer that is provided is by guess work or by actual thought process. This is done by presenting a feedback slide corresponding to each MCQ. The experimental procedure within each level of complexity is followed as shown in Fig. 5 and explained in Section III.B.

B. Participants

Thirteen final year MCA students pursuing interns (age groups: 23-25 years, 7 - females, 4 - males) are recruited for the study. All the subjects are right handed and had normal or corrected to normal vision with the aid of glasses. They all hail from similar cultural backgrounds. The experiment is carried out in a closed air-conditioned room to avoid any external distractions or noise. These factors are basically taken into consideration to do away with the variance in brain lateralization among the participants. A pre-experiment session is conducted for the participants wherein the protocols are explained. The main objective of determining stress against various stimuli is not told, since we suspect that this would bias the participants towards the stimulus. An informed consent form is signed by all the participants before the experiment. The data collected are anonymized and the results of the study do not reveal any personally identifiable information.

Among the thirteen participants, we have rejected participant or subject number *S2*, as the subject was electro-dermally stabile [24]. Apart from this, we have rejected the subjects *S1*, *S4*, *S7*, *S8* and *S9* from GSR analysis. The reason behind this rejection is given in the next section through detailed inspection of the GSR signals. Thus finally 7 subjects' data is used for GSR analysis. For HRV analysis, out of 13 subjects we have computed HRV for 11 subjects as *S2* is excluded from the experiment due to stabile nature and the data for *S9* is not logged properly due to error in data capture.

[3] http://www.brainmetrix.com/free-iq-test/

C. Inspection of raw GSR signals

The raw GSR signals are inspected for the 12 participants or subjects (except *S2*). It was noticed that for subject *S1* and *S9*, the GSR data is not logged for one of the stimuli. This is due to the error in running the Shimmer application during the experiment.

At the beginning of each level of stimulus there is a two minute "eye closed" rest phase. During this time, the GSR signal (skin conductance) is expected to reduce with time. A sample GSR signal during the "eye closed" rest phase is shown in Fig. 7. The sampling rate of the GSR signal is 10.2 Hz. Hence the end of rest phase of 120 seconds occurs at around 1224th sample and is marked with a vertical line.

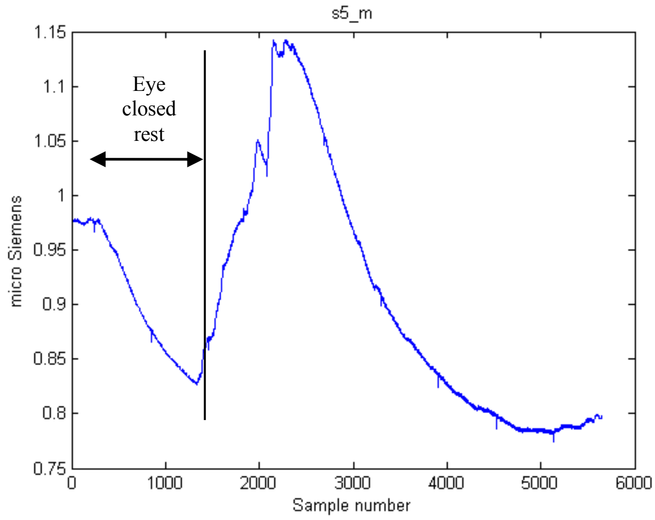

Fig. 7. Sample raw GSR signal (in microSiemens) for S5 moderate stimulus indicating the gradual decay of the signal during the eye closed rest state

For two subjects (*S4* and *S7*) the GSR signal did not reflect the initial rest condition of the subjects for the easy stimulus. Fig. 8 shows one of the raw GSR waveform where the expected decay in the rest state is not visible.

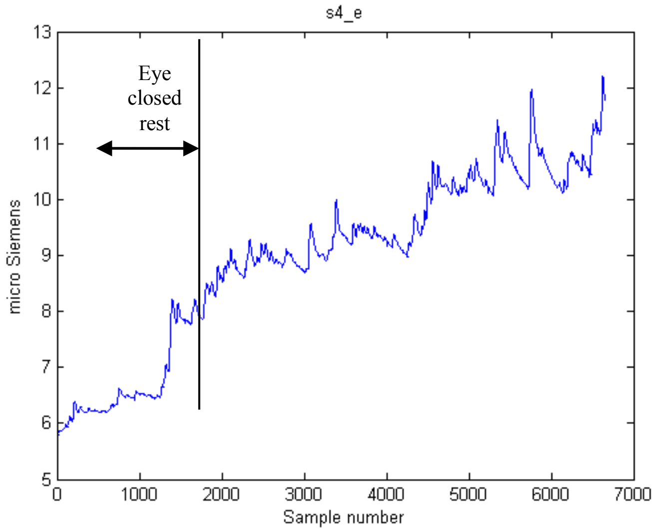

Fig. 8. Sample raw GSR signal (in microSiemens) for S4 easy stimulus not having the gradual decay of the signal during the eye closed rest state

These two subjects belong to two different groups hence one of them had easy stimulus at the first and the other had the same as third stimulus. Possible reasons for this may be that the subject did not follow the instruction properly or might have closed the eyes but could not mentally rest due to the anxiety of the experiment or for some other unknown reason. This particular issue needs to be investigated in future.

The data for one more subject (*S8*) is rejected as the characteristics of the GSR signal is not observed and a continuous decay of the GSR signal is seen as shown in Fig. 9. This is due to loose contact of the electrodes with the skin. In future, this needs to be automatically detected through signal processing before the start of the actual experiment.

Thus we process with the remaining 7 subjects' data for further analysis.

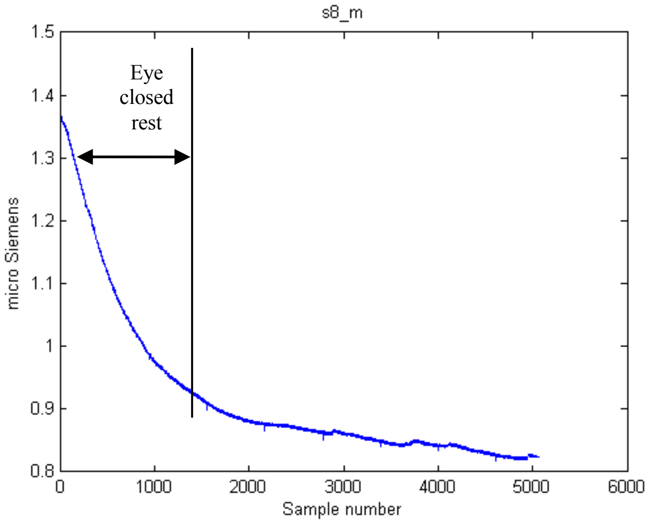

Fig. 9. Sample raw GSR signal (in microSiemens) for S8 moderate stimulus where the decay continues beyond the eye closed reset state, indicating a loose contact

D. Results of GSR analysis

For the completeness of understanding, initially a good sample raw GSR signal (for *S10* difficult stimulus) is shown in Fig. 10. It can be seen that after the initial decay in the eye closed rest state, once the stimulus starts then the GSR signal reflects the change in skin conductance due to a task.

The GSR data is subdivided into a number of windows of duration 10 seconds with 50% overlap between the windows. Among multiple features described in Section III.C, we present the results for the score generated from the fluctuation analysis as it outperforms the tonic / phasic power and peak detection features. This is due to the fact that the fluctuation analysis inherently combines both the tonic / phasic information and also the variation of the peaks in the raw GSR signal. TABLE I indicates the score computed using (7) with the feature extracted by fluctuation analysis (6). Out of 7 subjects, for 4 subjects the sequence of MCQs are *easy*, *moderate* and *difficult*; while for the remaining 3 it is *difficult*, *moderate* and *easy*. It can be seen that except *S3*, for all the subjects, the score increases with the increase in the complexity level. This indicates that score computed from the GSR signal reflects the mental load or stress in an individual while they are answering the MCQs. This also indicates that we would be able to generate an expected stress score for a given set of MCQs through this type of control study and then use the same as an additional attribute for the MCQ bank.

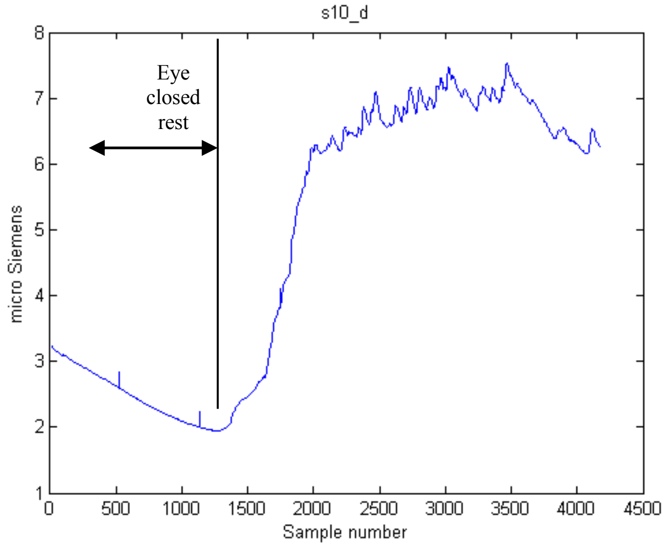

Fig. 10. Sample raw GSR signal (in microSiemens) for S10 difficult stimulus where there is a decay in the eye closed rest state and the variation in GSR signal during performing the task

TABLE I. SCORE GENERATED USING FLUCTUATION ANALYSIS FEATURE FOR THREE TYPES OF STIMULUS (MCQS)

Subjects	Easy	Moderate	Difficult
S3	0.951276	0.856829	0.868389
S5	0.782963	0.82167	0.8769934
S6	0.793012	0.81899	0.8259064
S10	0.552243	0.719663	0.9755355
S11	0.887408	0.90616	0.9368579
S12	0.509725	0.492873	0.6361562
S13	0.827506	0.741369	0.8894668

E. Results of heart rate variability analysis

PPG signal gives the indication of instantaneous blood flow through blood vessels. Initially the raw signal is normalized and shifted to zero mean. Subsequently a band pass filter with cut-off 0.5 Hz and 5 Hz is used to remove the unwanted high frequency components. Finally, a peak detection is used to calculate the n-n interval and then the HRV is calculated using SDNN [28]. Out of 13 participants or subjects we have computed HRV for 11 subjects as *S2* and *S9* are excluded as explained earlier.

The gold standard for HRV computation is by using the process of Electrocardiography (ECG) [28]. The n-n intervals computed from PPG provides a good estimate of the HRV, though inferior compared to the one obtained from ECG [30]. Due to this reason we have limited our analysis by comparing only the two extremities of the complexities namely *easy* and *difficult*. It is to be noted that for the unobtrusiveness nature of obtaining SPO2 compared to ECG, we have used the PPG waveform which is good enough for our initial analysis. Results are presented for SDNN as it is the most important time domain parameter [29]. A comparison of the SDNN computed for *easy* and *difficult* MCQs is shown in Fig. 11.

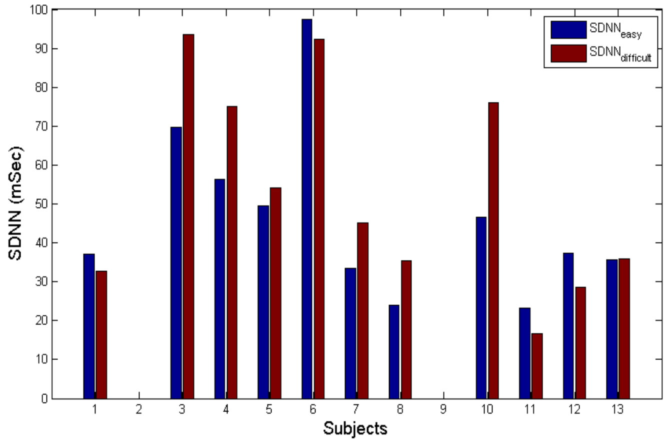

Fig. 11. Bar plot for the SDNN (mSec) computed for *easy* ($SDNN_{easy}$) and *difficult* ($SDNN_{difficult}$) MCQs

With the increase in the difficulty level, it is expected that SDNN would decrease as the parasympathetic nervous system takes control and lowers the HRV [29]. This behavior is seen only for few subjects namely *S1*, *S6*, *S11*, *S12* and hence demands further investigation. While exploring the reason behind the same, we analyzed the amount of MCQs answered by the subject through guess work. TABLE II shows the percentage of MCQs answered by guess where we can observe large amount of guess work. This indicates a possibility of the subjects not experiencing the stress during the experiment.

TABLE II. PERCENTAGE (%) OF MCQS ANSWERED BY GUESS FOR THREE TYPES OF STIMULUS (MCQS)

Subjects	Easy	Moderate	Difficult
S1	33	37.5	20
S3	50	25	50
S4	25	12.5	33
S5	8.3	37.5	16.7
S6	41.7	37.5	33
S7	33	12.5	50
S8	33	37.5	33
S10	8.3	12.5	16.7
S11	33	12.5	16.7
S12	33	12.5	0
S13	100	75	12.5

Next we investigate for a possible relation between the guess and the change in HRV. It is expected that lower the guess work more the experienced mental load. Hence we compute the difference $(Guess_{diff})$ in the percentage guess between the *difficult* and the *easy* MCQs as shown in (8).

$$Guess_{diff} = Guess_{difficult} - Guess_{easy} \quad (8)$$

The lower the values of $Guess_{diff}$ more is the experienced relative stress for the *difficult* MCQs compared to the *easy* ones and hence we expect the relative HRV to be lower for the *difficult* MCQs. Thus we compute the difference of SDNN values $(SDNN_{diff})$ for *easy* and *difficult* MCQs as shown in (9).

$$SDNN_{diff} = SDNN_{difficult} - SDNN_{easy} \quad (9)$$

Fig. 12 shows the bar plot comparing the relation between $Guess_{diff}$ and $SDNN_{diff}$. It can be seen that for most of the subjects they follow the same trend. This indicates that the guess work has an effect in the amount of reduction of HRV. A negative value of $SDNN_{diff}$ indicates the desired behavior where a subject experiences higher stress for *difficult* MCQs compared to *easy* ones. For subjects *S3* and *S8*, the percentage guess in *easy* and *difficult* are same and hence the bar plots for $Guess_{diff}$ do not show up for those subjects. For *S3*, the percentage guess is 50% which is higher compared to 33% for S8. This observation is also in relation with the $SDNN_{diff}$, where *S3* has higher $SDNN_{diff}$ than that of *S8*. Subject S13 guessed all the answers in *easy* MCQ and hence is treated as an outlier.

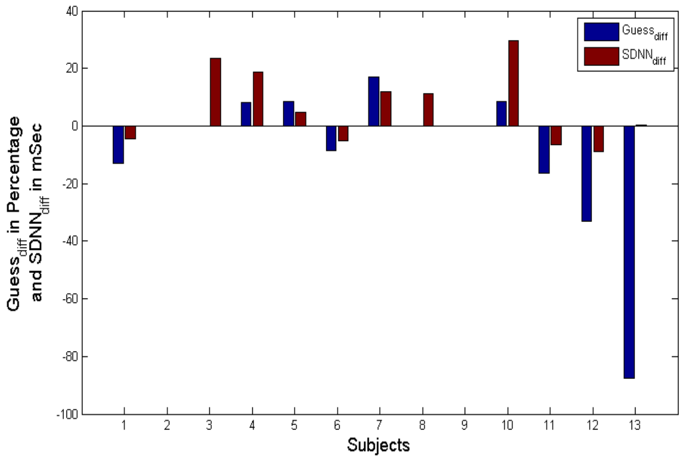

Fig. 12. Bar plot for analyzing the relation between the $SDNN_{diff}$ (mSec) and $Guess_{diff}$ (Percentage)

It is to be noted that the standard method to compute HRV [29] requires at least 1 minute of PPG data and larger the duration, better is the estimate of HRV. The duration of the stimulus in our experiment ranges from 3 to 5 minutes and hence a single value of HRV is computed for each level of complexity of MCQ. Whereas in case of the GSR the feature values for the fluctuation analysis is computed for every 10 seconds window. This may be the reason for the score computed from GSR is not effected by the guess work whereas the observed HRV has an effect. Another major difference between the two sensors is the sampling rate, for GSR it is 10.2 Hz and for SPO2 it is 60 Hz. This may be a reason that the measured HRV reflects certain information on the guess work.

V. CONCLUSION & FUTURE SCOPE

In this paper, we have tried to measure the mental stress from certain physiological signals like heart rate measurement and galvanic skin response during assessment of a student. MCQs are designed and grouped into three levels of complexities based on expert ratings. Participants are chosen such that the IQ levels ranges from low to high. Each level of MCQs are presented to the participants after an initial closed eye rest state. The sequence of complexity is evenly distributed to avoid any biasing effect. Results show that the score obtained from the fluctuations analysis of the GSR signals correlate well with the complexity of the MCQs. The HRV analysis gives valuable insights into the guess work done while answering the MCQs.

In future, the length and depth of the interleaving for the complexity levels of the MCQs would be varied to understand their effect in the stress level. Effort would be directed to formulate a model based on the relation between the stress level, length and depth of the interleaved sequence of the MCQs and the student profile. Moreover, to understand the correlation properties of the signal, short, medium and long term correlation analysis would be done by using multiple α of the Detrended Fluctuation Analysis. In regards to GSR data analysis, in future, we would like to focus on automatic detection and rejection of labile and stabile participants, automatic detection of erroneous signals due to loose contact of the sensors with the skin and wireless communication errors. A detailed study would be done to understand the effect of the sampling rate of the sensors on the reflection of stress level and the extent of guess work during an examination process.

REFERENCES

[1] Hart, Sandra G., and Lowell E. Staveland. "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research." *Advances in psychology* 52 (1988): 139-183.

[2] Paas, Fred GWC, Jeroen JG Van Merriënboer, and Jos J. Adam. "Measurement of cognitive load in instructional research." *Perceptual and motor skills* 79.1 (1994): 419-430.

[3] Bratfisch, Oswald. "Perceived Item-Difficulty in Three Tests of Intellectual Performance Capacity." (1972).

[4] Csíkszentmihályi, Mihály, "Flow: The Psychology of Optimal Experience", New York: Harper and Row, 1990, ISBN 0-06-092043-2

[5] Csikszentmihalyi, Mihaly, and Isabella Selega Csikszentmihalyi, eds. Optimal experience: Psychological studies of flow in consciousness. Cambridge University Press, 1992

[6] Novak, Thomas P., and Donna L. Hoffman. "Measuring the flow experience among web users." Interval Research Corporation 31 (1997)

[7] Nakamura, J. and Csíkszentmihályi, M. Flow Theory and Research. In Snyder, C. R. and Lopez, S. J. eds. Oxford handbook of Positive Psychology.Oxford University Press, Oxford, 2009, 195-206

[8] Haapalainen, Eija, et al. "Psycho-physiological measures for assessing cognitive load." *Proceedings of the 12th ACM international conference on Ubiquitous computing*. ACM, 2010.

[9] Navalyal, Geeta U., and Rahul D. Gavas. "A dynamic attention assessment and enhancement tool using computer graphics." *Human-centric Computing and Information Sciences* 4.1 (2014): 1-7.

[10] Sinha, Aniruddha, et al. "Dynamic assessment of learners' mental state for an improved learning experience." *Frontiers in Education Conference (FIE), 2015. 32614 2015. IEEE*. IEEE, 2015.

[11] Critchley E (2002) Electrodermal responses: what happens in the brain. Neuroscientist 8(2):132–142

[12] Vrijkotte, Tanja GM, Lorenz JP Van Doornen, and Eco JC De Geus. "Effects of work stress on ambulatory blood pressure, heart rate, and heart rate variability." Hypertension 35.4 (2000): 880-886.

[13] Healey, Jennifer, and Rosalind Picard. "SmartCar: detecting driver stress." Pattern Recognition, 2000. Proceedings. 15th International Conference on. Vol. 4. IEEE, 2000.

[14] J. Healey and R. W. Picard. Startlecam: A cybernetic wearable camera. In ISWC, pages 42-49, 1998

[15] Tarvainen, M. P., et al. "Principal component analysis of galvanic skin responses." Engineering in Medicine and Biology Society, 2000. Proceedings of the 22nd Annual International Conference of the IEEE. Vol. 4. IEEE, 2000.

[16] Benoit, Alexandre, et al. "Multimodal focus attention and stress detection and feedback in an augmented driver simulator." Personal and Ubiquitous Computing 13.1 (2009): 33-41.

[17] Zhai, Jing, et al. "Realization of stress detection using psychophysiological signals for improvement of human-computer interactions." SoutheastCon, 2005. Proceedings. IEEE. IEEE, 2005.

[18] de Santos Sierra, Alberto, et al. "Two stress detection schemes based on physiological signals for real-time applications." Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2010 Sixth International Conference on. IEEE, 2010.

[19] Ahuja, Nutan D., et al. "GSR and HRV: its application in clinical diagnosis." Computer-Based Medical Systems, 2003. Proceedings. 16th IEEE Symposium. IEEE, 2003.

[20] Lin, Tao, et al. "Do physiological data relate to traditional usability indexes?." Proceedings of the 17th Australia conference on Computer-Human Interaction: Citizens Online: Considerations for Today and the Future. Computer-Human Interaction Special Interest Group (CHISIG) of Australia, 2005.

[21] Nourbakhsh, Nargess, et al. "Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks." Proceedings of the 24th Australian Computer-Human Interaction Conference. ACM, 2012.

[22] Zhou, Jianlong, Ju Young Jung, and Fang Chen. "Dynamic Workload Adjustments in Human-Machine Systems Based on GSR Features." Human-Computer Interaction–INTERACT 2015. Springer International Publishing, 2015. 550-558.

[23] Garcia-Mancilla, Jesus, and Victor M. Gonzalez. "Stress Quantification Using a Wearable Device for Daily Feedback to Improve Stress Management." Smart Health. Springer International Publishing, 2015. 204-209.

[24] Dawson, M E., et al."The Electrodermal System."Handbook of psychophysiology (2007).

[25] Bakker, Jorn, Mykola Pechenizkiy, and Natalia Sidorova. "What's your current stress level? Detection of stress patterns from GSR sensor data." Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on. IEEE, 2011.

[26] Braithwaite, Jason J., et al. "A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments." Psychophysiology 49 (2013): 1017-1034.

[27] J. Taelman, S. Vandeput, A. Spaepen and S. Van Huffel, "Influence of Mental Stress on Heart Rate and Heart Rate Variability", 4th European Conference of the International Federation for Medical and Biological Engineering, vol. 22, pp. 1366-1369, 2009

[28] Malik, M. Heart Rate Variability: Standards of measurement, physiological interpretation, and clinical use. Ann. of Noninvasive Electrocardiology 1, 2, 151–181, 1996

[29] Acharya, U.R., Joseph, K.P., Kannathal, N., Lim, C.M., and Suri, J.S. Heart rate variability: A review. Med & Bio Engin & Comp 44, 12, 1031–1051, 2006

[30] Selvaraj, Nandakumar, et al. "Assessment of heart rate variability derived from finger-tip photoplethysmography as compared to electrocardiography." Journal of medical engineering & technology 32.6 (2008): 479-484.